Performance Analysis of Machine Learning based Models for Effective Prediction of Cardiovascular Diseases

Synopsis submitted to the IIMT University in partial fulfillment of the requirements for the award of the Degree of

DOCTOR OF PHILOSOPHY

In

Electrical and Electronics Engineering

By

(Dinesh Kumar Yadav)
Roll No: 202107075
Admission No. 2162336

Part time

Under the guidance of

Dr. Ankur Gupta
School of Engineering & Technology
IIMT University, Meerut
Co-Guide: Prof. (Dr.) Seema Nayak
IIMT College of Engineering, Greater Noida

IIMT University, Meerut

(School of Engineering & Technology)
(Feb 2023)

1. Introduction:

One of the most common reasons of death all across the world is cardiovascular disease. It is a prominent health issue. But at the same time their lives can be saved by early identification of cardiovascular disorders such as heart attacks and heart strokes. A majority of fatalities from cardiac disorders are due to the delayed detection of the A small percentage of fatalities from cardiac disorders are caused by natural clinical causes, while the majority one is caused by the delayed detection of the disease. It requires the implementation of some accurate, reliable, and practical approaches for early detection of the disease. In this regards in the field of healthcare, two popular approaches can be made use of for examining and evaluating huge datasets namely machine learning and deep learning algorithms. A lot of experts and researchers have in the past, practiced a range of both deep learning and machine learning approaches for examining and evaluating huge quantities of medical data, which assist the doctors in predicting heart problems. At present, there is a lot of data by a large number of doctor's facilities and also by social insurance sectors which collect a patient's information which is generally hard to deal with as of currently existing frameworks. [1]

The aim of this research is to accurately predicting if any patient has the symptoms of cardiovascular diseases or not. We have studied and implemented many traditional machine learning algorithms like Logistic Regression (LR), Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boost (XG-Boost) and K-Nearest Neighbor algorithm (KNN). The output of all these models will be evaluated and then implement a model with higher accuracy.

2. Literature Review:

There is a lot of data by a large number of doctor's facilities and also by social insurance sectors which collect a patient's information which is generally hard to deal with as of presently existing frameworks [1]. In [2] a hybrid model was implemented by the authors for prediction of cardiac disease. The authors used a model which combines the two machine learning methods namely the random forest and the decision tree methods. These were used on the Cleveland dataset. A total of 303 cases are there in this collection, with having 14 different features. In this research an attempt was made to use machine learning and automated diagnosis system to predict the cardiac disease. The accuracy obtained by DT and RF algorithms was found to be 79% and 81% respectively. The authors of [3] implemented the Decision Tree, KNN and the Random Forest machine learning algorithms and RF techniques and with this they obtained 100% accuracy. They also observed that the supervised ML algorithms are applicable in order to create a precise cardiovascular disease prediction model with a range of applications.

Further, in [4] a supervised learning machine learning algorithm model was used like Naïve Bayes, decision trees, K-nearest neighbor and random forest algorithms. Here the dataset maintained by Irvine's University of California's Cleveland database was used which is a dataset of the heart disease patients. This dataset consisted of total 303 incidences and a total of 76 features. The purpose of their study was to predict the chance of a person developing cardiac

issues. In [5] the authors performed review and presented some latest work in the machine learning area and also image-fusion involving the classification methods. The authors found that the ANN has a very significant effect in predicting a heart disease in most of the models. [8] presumed that although most of the analysts utilize varied classifier methods, like SVM, KNN and Neural system, applying Naïve Bayes and Decision tree provides a better outcome in the finding of coronary related issues and a lot better exactness when compared with other classifiers. [9] used real data of patients taken from hospital to test prediction model. Both Structured and unstructured data was obtained to clean the missing and noisy data. After this the dataset was classified into training and test dataset and different ML algorithms are applied namely CNN, Naïve Bayes and KNN and others. The accuracy which was found using these are 85%, 80% and 74% respectively. The newly created model achieved an accuracy of approximately 85 to 88%. The accuracy of the model is highest for CNN model, closely followed by Naive Bayes method, which is somewhat less accurate than CNN, and finally KNN algorithm, which was found to have the least accuracy.

A demonstration is made in [10] on the prediction of survival rates and also for the mistake rates. It makes an attempt to analyze the death rate. [11] gives the exactness among the various calculations by discovering the death rate.

3. Objective of Work:

- To perform the research literature review of previous work done in the area.
- To develop an algorithm by utilizing a training dataset to train the machine learning algorithm and also predicts cardiac attack by using the input data provided additionally.
- To analyze the different machine learning based models for effective prediction of cardiovascular diseases

4. Research Methodology / Work Plan:

To analyze the different machine learning based models for effective prediction of cardiovascular diseases, we shall create the system which can successfully learn from its previous experiences and also make predictions [6]. It develops an algorithm by utilizing a training dataset to train the machine learning algorithm and also predicts cardiac attack by using the input data provided additionally. The model is created by using machine learning by assessing the given dataset and discovering the unseen patterns from it, and for the new datasets it can correctly predict the outcomes correctly.

Finally by using the above defined updated information and input analysis for the purpose of predicting potential cardiovascular disease an algorithm is then tested for its accuracy [4].

5. Conclusion:

Cardiovascular disease is one of the most common causes of death amongst all of the causes of mortality. Due to limitations and lack of knowledge and correct warning indicators of heart

failure, it is difficult to detect a cardiovascular issue. We are going to use different machine learning algorithms in this research work and then create an algorithm which will have more accuracy than the presently available algorithms. Also, we shall collect real time data from hospitals instead of readymade data-set available.

6. Bibliography:

1. Shankar, V., Kumar, V., Devagade, U., Karanth, V., & Rohitaksha, K. (2020). Heart disease prediction using CNN algorithm. *SN Computer Science*, *1*(3), 170.

2. Devi, R. K., & Muthukannan, M. (2021, January). Self-healing fault tolerance technique in cloud datacenter. In *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (pp. 731-737). IEEE.

3. Desai, U., & Mantri, S. (2022). Hybrid Model of Machine Learning Algorithms for Prediction of Cardiovascular Disease. *Journal of Positive School Psychology*, 10551-10560.

4. Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, *1*, 1-6.

5. Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., & Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. *Materials Today: Proceedings*, *37*, 3213-3218.

6. Hossen, M. A., Tazin, T., Khan, S., Alam, E., Sojib, H. A., Monirujjaman Khan, M., & Alsufyani, A. (2021). Supervised machine learning-based cardiovascular disease analysis and prediction. *Mathematical Problems in Engineering*, *2021*, 1-10.

7. Dinesh, K. G., Arumugaraj, K., Santhosh, K. D., & Mareeswari, V. (2018, March). Prediction of cardiovascular disease using machine learning algorithms. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)* (pp. 1-7). IEEE.

8. Parthiban, G., & Srivatsa, S. K. (2012). Applying machine learning methods in diagnosing heart disease for diabetic patients. *International Journal of Applied Information Systems*, *3*(7), 25-30.

9. Kaur, B., & Singh, W. (2014). Review on heart disease prediction system using data mining techniques. *International journal on recent and innovation trends in computing and communication*, *2*(10), 3003-3008.

10. Rubini, P. E., Subasini, C. A., Katharine, A. V., Kumaresan, V., Kumar, S. G., & Nithya, T. M. (2021). A cardiovascular disease prediction using machine learning algorithms. *Annals of the Romanian Society for Cell Biology*, 904-912.

11. Sana Bharti, Shailendra Narayan Singh" Analytical study of heart disease comparing with different algorithms": Computing, Communication & Automation (ICCCA), 2015 International Conference.